



**UNIVERSITI PUTRA MALAYSIA**

**A STRUCTURAL STUDY ON THE SPECIFICITY OF F1 PROTEASE**

**AZIRA MUHAMAD.**

**FBSB 2005 34**

# **A STRUCTURAL STUDY ON THE SPECIFICITY OF F1 PROTEASE**

**By**

**AZIRA MUHAMAD**

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in  
Fulfilment of the Requirements for the Degree of Master of Science**

**December 2005**



Abstract of the thesis presented to the Senate of Universiti Putra Malaysia in  
fulfilment of the requirements for the degree of Master of Science

**A STRUCTURAL STUDY ON THE SPECIFICITY OF F1 PROTEASE**

By

**AZIRA MUHAMAD**

**December 2005**

**Chairman : Assoc. Professor Raja Noor Zaliha Raja Abd. Rahman, PhD**

**Faculty : Biotechnology and Biomolecular Sciences**

Specificity studies of a thermostable alkaline serine protease F1 with its substrates were carried out through computational docking method. Structures of a series of synthetic peptide substrates were docked to the active site of the homology modelled F1 protease using AutoDock 3.0.5. The resulting clusters of the substrates that were docked were analysed by inspecting the energetic results and the orientation of each cluster to determine the arrangement of productive binding. The amino acids of the binding site that participated in the hydrophobic and hydrogen-bond interactions were also determined. Docking results showed that all substrates tested bound near the catalytic residues with SucAAPFpNA, the biggest substrate, showing the most negative docked energy value ( $E_{\text{docked}} = -18.75$  kcal/mol). Smaller substrates such as GpNA and AApNA showed higher docked energy ( $E_{\text{docked}} = -7.77$  kcal/mol and  $-8.77$  kcal/mol, respectively). The

best docked structure of each substrate was determined from the clusters. It was found that most of the lowest  $E_{\text{docked}}$  conformations display the best docked orientations with respect to the least distance calculated between the carbonyl carbon of the substrate P1 residue and  $\gamma$ -oxygen of the Ser226 catalytic triad. From the results, it also demonstrated that S1, S2 and S4 subsites of the enzyme play a critical role in determining the substrate specificity of F1 protease from the point of view that bigger-sized substrates such as SucAAPFpNA and SucAAPLpNA showed more favourable  $E_{\text{docked}}$ . This work also support the hypothesis that the catalytic serine and histidine residues were essential in catalysis as well as in stabilizing the enzyme-substrate complex for binding.

Validation of computational study was carried out through biochemical assay. It was found that SucAAPFpNA was the most preferred substrate for the enzyme with specific activity of 3.079 U/mg followed by SucAAPLpNA at 1.016 U/mg. SucAAPFpNA was also observed to show the highest binding affinity towards the protease ( $K_m = 1.26\text{mM}$ ) and the highest catalytic ratio ( $1.226 \text{ min}^{-1}.\text{mM}^{-1}$ ) compared to the other substrates tested. Similar to computational observations, smaller peptides showed lower specific activity and binding affinity towards the protease. Rank-order of the substrates tested for the docking and experimental methods were found to be similar for the top two substrates, with lesser agreement for the other substrates.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia  
sebagai memenuhi keperluan ijazah Master Sains

## **KAJIAN STRUKTUR SPESIFIKASI F1 PROTEASE**

Oleh

**AZIRA MUHAMAD**

**Disember 2005**

**Pengerusi : Profesor Madya Raja Noor Zaliha Raja Abd. Rahman, PhD**

**Fakulti : Bioteknologi dan Sains Biomolekul**

Suatu penyelidikan spesifikasi terhadap serine protease F1 yang stabil suhu tinggi dan bersifat alkali bersama substrat telah dijalankan melalui kaedah percantuman pengkomputeran. Struktur-struktur peptida sintetik telah dicantum ke tapak aktif F1 protease yang telah dimodelkan secara homologi dengan menggunakan AutoDock 3.0.5. Kelompok yang telah dihasilkan oleh substrat-substrat yang telah dicantum, telah dianalisa keputusan tenaga dan orientasi setiap kelompok untuk menentukan susunan cantuman yang produktif. Asid amino di tapak pengikat yang menyertai interaksi hidrofobik dan ikatan hidrogen juga telah ditentukan. Percantuman substrat-substrat pada enzim ini menunjukkan yang kesemua substrat yang telah diuji mengikat berhampiran residu katalitik di mana SucAAPFpNA, iaitu substrat yang paling besar, menunjukkan nilai  $E_{\text{docked}}$  yang paling rendah, -18.75 kcal/mol. Substrat-substrat yang lebih kecil seperti GpNA dan AAPNA menunjukkan nilai  $E_{\text{docked}}$  yang lebih tinggi, masing-masing pada -7.77

kcal/mol dan -8.77kcal/mol. Struktur yang paling sesuai telah ditentukan daripada setiap kelompok. Didapati bahawa kebanyakan konformasi  $E_{\text{docked}}$  yang paling rendah mempunyai orientasi yang paling sesuai berpunca dari jarak di antara karbon karbonil substrat residu P1 dan  $\gamma$ -oksigen Ser226 triad katalitik yang paling rendah. Cantuman substrat-substrat itu juga menunjukkan subtapak S1, S2 dan S4 enzim memainkan peranan yang kritikal dalam penentuan spesifikasi substrat bagi F1 protease kerana didapati bahawa substrat-substrat yang lebih besar seperti SucAAPFpNA dan SucAAPLpNA menunjukkan nilai  $E_{\text{docked}}$  yang lebih rendah. Kajian ini mengesahkan bahawa residu-residu katalitik seperti serine dan histidine adalah penting dalam katalisis dan dalam pengstabilan kompleks enzim-substrat untuk percantuman.

Kajian kesahihan telah dijalankan dengan melakukan asai biokimia. Didapati bahawa SucAAPFpNA adalah substrat yang paling digemari oleh enzim ini dengan aktiviti spesifik, 3.079 U/mg, diikuti SucAAPLpNA pada 1.016 U/mg. SucAAPFpNA juga didapati mempunyai afiniti pengikatan yang paling tinggi terhadap protease ( $K_m = 1.26\text{mM}$ ) dan ratio katalitik paling tinggi ( $1.226 \text{ min}^{-1}.\text{mM}^{-1}$ ) berbanding dengan substrat-substrat lain yang diuji. Peptida yang lebih kecil didapati mempunyai spesifik aktiviti dan afiniti pengikatan yang rendah terhadap F1 protease, seperti yang telah diperolehi dari analisis pengkomputeran. Persamaan yang didapati daripada urutan substrat-substrat yang diperolehi daripada percantuman pengkomputeran

dan kaedah eksperimental adalah pada tempat kedua teratas, substrat-substrat lain kurang menunjukkan kaitan.

## ACKNOWLEDGMENT

I wish to express my deepest gratitude to my supervisor, Assoc. Prof. Dr. Raja Noor Zaliha Raja Abd. Rahman for her constant guidance and encouragement throughout the duration of this study. My sincere appreciation is extended to my co-supervisors, Prof. Dr. Abu Bakar Salleh, Prof. Dr. Mahiran Basri and Dr. Habibah A. Wahab. Their critical comments and support have been an invaluable source of motivation. My gratitude is also extended to Assoc. Prof. Dr. Mohd. Basyaruddin Abd. Rahman for his helpful comments and guidance.

I would like to thank all my lab mates, Ferrol, Kak Ain, Kak Ina, Kak Naz, Leow, Ghanie, Rofandi, Kak Lia, Kok Whye, Brother Mohd, Aiman, Sha, Shuk, Sue, Yob, Ada, Chee Fah and Tengku for their friendship and support during my stay at the lab. You guys have been really helpful. My sincere appreciation also goes especially to Bimo for his invaluable encouragement and comments. You have been a real source of inspiration. Thank you for being so patient and thank you so much for the long hours you've spent for me.

Last but not least, my heartfelt gratitude goes to my family, Bapak, Mak, Abang, Kak Yani and Alen. Thank you for your patience, encouragement and support.





I certify that an Examination Committee met on 14<sup>th</sup> December 2005 to conduct the final examination of Azira Muhamad on her Master of Science thesis entitled “A Structural Study of the Specificity of F1 Protease from *Bacillus stearothermophilus*” in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

**Mohd Arif Syed, PhD**

Professor

Faculty of Biotechnology and Biomolecular Sciences

Universiti Putra Malaysia

(Chairman)

**Suhaimi Napis, PhD**

Associate Professor

Faculty of Biotechnology and Biomolecular Sciences

Universiti Putra Malaysia

(Internal Examiner)

**Nor Aripin Shamaan, PhD**

Professor

Faculty of Biotechnology and Biomolecular Sciences

Universiti Putra Malaysia

(Internal Examiner)

**Sharifuddin M Zain, PhD**

Associate Professor

Faculty of Science

Universiti Malaya

(External Examiner)



---

**HASANAH MOHD. GHAZALI, PhD**

Professor/Deputy Dean

School of Graduate Studies

Universiti Putra Malaysia

Date: **19 JAN 2006**

This thesis submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee are as follows:

**RAJA NOOR ZALIHA RAJA ABD. RAHMAN, PhD**

Associate Professor

Faculty of Biotechnology and Biomolecular Sciences

Universiti Putra Malaysia

(Chairman)

**ABU BAKAR SALLEH, PhD**

Professor

Faculty of Biotechnology and Biomolecular Sciences

Universiti Putra Malaysia

(Member)

**MAHIRAN BASRI, PhD**

Professor

Faculty of Science

Universiti Putra Malaysia

(Member)

**HABIBAH A. WAHAB, PhD**

Lecturer

School of Pharmaceutical Sciences

Universiti Sains Malaysia

(Member)



---

**AINI IDERIS, PhD**

Professor/Dean

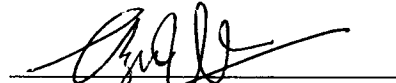
School of Graduate Studies

Universiti Putra Malaysia

Date: **07 FEB 2006**

## DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations, which have been duly acknowledge. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

  
AZIRA MUHAMAD

Date: 30.12.2005

## TABLE OF CONTENTS

	Page
ABSTRACT	ii
ABSTRAK	iv
ACKNOWLEDGMENTS	vii
APPROVAL	viii
DECLARATION	x
LIST OF TABLES	xiii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS	xix
 CHAPTER	
 1 INTRODUCTION	 1
 2 LITERATURE REVIEW	 4
2.1 Protease	4
2.1.1 Subtilisin-like serine protease	5
2.1.2 Structural properties of serine protease	8
Substrate Specificity	11
Theoretical Approaches	14
2.1.3 Thermostable protease from <i>Bacillus</i>	16
<i>stearothermophilus</i> F1	
2.1.4 Purification of proteases	17
2.1.5 Kinetic studies of proteases	18
2.1.6 Applications of proteases	20
2.2 Molecular Modelling	22
2.2.1 Docking	24
Docking Programs	25
Applications of docking	27
 3 MATERIALS AND METHODS	 29
3.1 Chemicals	29
3.2 Computer Hardwares and Softwares	30
3.3 Computer Simulations	31
3.3.1 Modelling of enzyme	31
3.3.2 Modelling of Peptides	31
3.3.3 Docking Calculations	37
3.3.4 Analysis	40
3.4 Experimental procedures	40
3.4.1 Screening for Positive Colonies	40
3.4.2 Assay of Protease Activity	40
3.4.4 Purification of <i>Bacillus stearothermophilus</i>	
protease F1	41
Crude Enzyme Preparation	41



	Heat Treatment	41
	Ultrafiltration	42
3.4.5	Substrate Specificity	42
	Specific Activity	42
	Kinetic Study	43
<b>4</b>	<b>RESULTS</b>	<b>44</b>
4.1	Computer Simulations	44
4.1.1	Modelling of enzyme	44
4.1.2	Flexible docking of substrates	44
	SucAAPFpNA	48
	SucAAPLpNA	51
	VLKpNA	51
	AAFpNA	54
	GltFpNA	59
	AceFpNA	59
	FpNA	62
	AApNA	67
	GpNA	70
4.2	Experimental Procedures	73
4.2.1	Purification of <i>Bacillus stearothermophilus</i> protease F1	73
4.2.2	Substrate Specificity	73
	Specific activity	73
	Kinetic study	75
<b>5</b>	<b>DISCUSSION</b>	<b>81</b>
5.1	Computer Simulations	81
5.1.1	Flexible docking of substrates	81
5.2	Experimental Procedures	85
5.2.1	Purification of <i>Bacillus stearothermophilus</i> protease F1	85
5.2.2	Substrate Specificity	86
5.3	Experimental Validation of Docking Results	90
<b>6</b>	<b>CONCLUSION AND RECOMMENDATION</b>	<b>97</b>
	<b>REFERENCES</b>	<b>99</b>
	<b>APPENDICES</b>	<b>107</b>
	<b>BIODATA OF THE AUTHOR</b>	<b>147</b>

## LIST OF TABLES

Table	Page
1 Substrates chosen for docking. The curly arrows indicate the rotatable bonds defined in the flexible docking. The bold residue is the P1 amino acid	34
2 Docking parameters	39
3 Results obtained for docking of various substrates to F1 protease. The numbers in brackets are the cluster rank of the conformation chosen	47
4 Purification table of recombinant F1 protease	74
5 Substrate specificity of F1 protease with peptide substrates	74
6 $K_m$ , $V_{max}$ , $k_{cat}$ and specificity of F1 protease with several substrates	80
7 The $\Delta G_{ES}$ calculated for the three substrates	80



## LIST OF FIGURES

Figure	Page
1 Serine protease mechanism (Voet and Voet, 1995)	7
2 Schematic representation of substrates/inhibitor binding to a subtilisin-like serine protease. Nomenclature P4- P2' and S4 - S2' according to Schechter and Berger (1967)	10
3 Three-dimensional structure of the predicted F1 protease. The residues rendered as sticks represent the catalytic triads	32
4 Active site of F1 protease. The enzyme is represented as atomic surfaces. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad.	45
5 Orientation of the lowest $E_{\text{docked}}$ of SucAAPFpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad.	49
6 Yellow lines indicate the distance of 3.88Å measured between the carbonyl carbon of the SucAAPFpNA P1 residue and $\gamma$ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type	49
7 Hydrogen-bond and hydrophobic interactions of the lowest $E_{\text{docked}}$ of SucAAPFpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds	50
8 Orientation of the lowest $E_{\text{docked}}$ of SucAAPLpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad	52

- 9 Yellow lines indicate the distance of 3.71Å measured between the carbonyl carbon of the SucAAPLpNA P1 residue and  $\gamma$ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type 52
  
- 10 Hydrogen-bond and hydrophobic interactions of the lowest  $E_{\text{docked}}$  of SucAAPLpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds 53
  
- 11 Orientation of the lowest  $E_{\text{docked}}$  of VLKpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad 55
  
- 12 Yellow lines indicate the distance of 5.11Å measured between the carbonyl carbon of the VLKpNA P1 residue and  $\gamma$ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type 55
  
- 13 Hydrogen-bond and hydrophobic interactions of the lowest  $E_{\text{docked}}$  of VLKpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds 56
  
- 14 Orientation of the lowest  $E_{\text{docked}}$  of AAFpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad 57
  
- 15 Yellow lines indicate the distance of 3.67Å measured between the carbonyl carbon of the AAFpNA P1 residue and  $\gamma$ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type 57
  
- 16 Hydrogen-bond and hydrophobic interactions of the lowest  $E_{\text{docked}}$  of AAFpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts



Dashed lines indicate the length of the hydrogen bonds	58
17 Orientation of the lowest $E_{\text{docked}}$ of GltFpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad.	60
18 Yellow lines indicate the distance of 3.86Å measured between the carbonyl carbon of the GltFpNA P1 residue and $\gamma$ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type	60
19 Hydrogen-bond and hydrophobic interactions of the lowest $E_{\text{docked}}$ of GltFpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds.	61
20 Interactions of AceFpNA with the active site of F1 protease (a) Orientation of the lowest $E_{\text{docked}}$ of AceFpNA docked to the active site of F1 protease. (b) Hydrogen-bond and hydrophobic interactions of the lowest $E_{\text{docked}}$ of AceFpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds.	63
21 Interactions of AceFpNA with the active site of F1 protease (a) Orientation of the best docked conformation of AceFpNA docked to the active site of F1 protease. (b) Hydrogen-bond and hydrophobic interactions of the best docked of AceFpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds.	64
22 Orientation of the lowest $E_{\text{docked}}$ of FpNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad.	65

- 23 Yellow lines indicate the distance of 5.12Å measured between the carbonyl carbon of the FpNA P1 residue and  $\gamma$ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type 65
- 24 Hydrogen-bond and hydrophobic interactions of the lowest  $E_{\text{docked}}$  of FpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds. 66
- 25 Orientation of the lowest  $E_{\text{docked}}$  of AAPNA docked to the active site of F1 protease. Pink coloured surface is the S1 subsite, cyan surface is S2 and S4 is the purple coloured surface while the blue surface represents the catalytic triad. 68
- 26 Yellow lines indicate the distance of 4.01Å measured between the carbonyl carbon of the AAPNA P1 residue and  $\gamma$ -oxygen of catalytic Ser226. The substrate and Ser226 are rendered as sticks and coloured according to the atom type 68
- 27 Hydrogen-bond and hydrophobic interactions of the lowest  $E_{\text{docked}}$  of AAPNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds. 69
- 28 Interactions of GpNA with the active site of F1 protease  
(a) Orientation of the lowest  $E_{\text{docked}}$  of GpNA docked to the active site of F1 protease. (b) Hydrogen-bond and hydrophobic interactions of the lowest  $E_{\text{docked}}$  of GpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length of the hydrogen bonds. 71
- 29 Interactions of GpNA with the active site of F1 protease  
(a) Orientation of the best docked structure of GpNA to the active site of F1 protease. (b) Hydrogen-bond and hydrophobic interactions of the best docked structure of GpNA with the active site of F1 protease represented by LigPlot. The comb-like arrangements indicate the hydrophobic contacts. Dashed lines indicate the length

of the hydrogen bonds	72
30 Time-course measurement of enzyme activity	76
31 Michaelis-Menten plot for the velocity data of SucAAPFpNA	76
32 Michaelis-Menten plot for the velocity data of SucAAPLpNA	77
33 Michaelis-Menten plot for the velocity data of AAPNA	77

## LIST OF ABBREVIATIONS

$\Delta G_{\text{binding}}$	Free energy of binding
$\text{\AA}$	Angstrom
dH <sub>2</sub> O	distilled water
$E_{\text{docked}}$	Docked energy
h	hour
HF	Hartree-Fock
K	Kelvin
kcal/mol	kilocalorie per mol
M	Molar
mg	milligram
ml	milliliter
mM	milliMolar
min	minute
RESP	restrained electrostatic potential
RMSd	root mean square deviation (s)
$\mu\text{g}$	microgram
$\mu\text{l}$	microliter
v/v	volume per volume
w/v	weight per volume

## CHAPTER 1

### INTRODUCTION

Proteases are enzymes that cleave peptide bonds at points within the protein and remove amino acids sequentially from either N or C-terminus. There are four mechanistic classes of proteases, they are serine proteases, cysteine proteases, aspartic proteases and metallo proteases. Two distinct families can be observed for serine proteases, which are the chymotrypsin family and the subtilisin family which include the bacterial enzymes such as subtilisin (Rao *et al.*, 1998). An example of subtilisin serine protease is *Bacillus stearothermophilus* protease F1. *B. stearothermophilus* was isolated from decomposed oil palm branches and its protease showing the characteristics of an extremely thermostable alkaline protease. *B. stearothermophilus* was shown to have a high degree of thermostability, able to grow up to 80°C with a pH range from 8.0 to 10.0. It has an optimum growth rate at 70°C and pH 9 (Rahman *et al.*, 1994).

Proteases differ in their properties such as substrate specificity, active site and catalytic mechanism. Structure-based mutational analysis of serine proteases specificity has produced a large database of information useful in addressing biological function and in establishing a basis for targeted design efforts (Perona and Craik, 1995). Despite extensive research on proteases, relatively little is known about the factors that control their

specificities. These questions that are difficult to answer experimentally, might be resolved theoretically.

Molecular modelling methods have been used in many ways to address problems in structural biology. The overall aim of modelling methods will often be to try to relate biological activity to structure (Forster, 2002). Molecular modelling is a general term that covers a wide range of molecular graphics and computational chemistry techniques used to display, build, manipulate, simulate and analyze molecular structure and to calculate properties of these structures. Through modelling, the understanding of properties, three-dimensional conformations, and enzymatic mechanism can also be improved.

One of the most useful areas of application of molecular modelling is the approach of fitting together, or docking, a protein to a second molecule. Molecular docking is important in understanding possible interactions between a protein and a ligand in the formation of a biologically important protein-ligand complex. It is also a suitable tool to estimate interaction energy of the binding step by means of rigid energy minimization methods (Teodoro *et al.*, 2001).

There are many docking software available and most of these programs are free for public. One of the widely used programs is AutoDock.

AutoDock is an automated docking program of flexible ligands to receptors. It was developed by Olson's group in 1990 (Goodsell and Olson, 1990). AutoDock uses free energy of the docking molecules using three-dimension potential-grids. The search algorithms implemented include Simulated Annealing (SA), Genetic Algorithm (GA) and Lamarckian GA (LGA) with GA + local search (LS) hybrid. The search and score method in AutoDock involves exploration of the configuration spaces available for interaction between ligand and receptor and it evaluates and ranks configurations using a scoring system which is the binding energy (Morris *et al.*, 1998).

In order to further understand the molecular basis of substrate specificity of *B. stearothermophilus* F1 protease, the following objectives were implemented in view of this research:

1. To find the best substrate based on binding energies obtained from docking procedure
2. To study the interaction of F1 protease with its substrates
3. To validate results obtained through computational studies by experimental methods

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 Proteases

Proteases comprise two groups of enzymes, the endopeptidases and the exopeptidases. The exopeptidases act only near the ends of polypeptide chains at the N or C terminus while the endopeptidases act preferentially in the inner regions of peptide chains away from the N and C termini. These proteases are classified according to their catalytic mechanisms. There are four mechanistic classes recognized by International Union of Biochemistry and Molecular Biology. They are serine proteases, cysteine proteases, aspartic proteases and metallo proteases (Rao *et al.*, 1998).

Serine proteases have been grouped into six distinct families of which the chymotrypsin-like clan and the subtilisin-like clan are the two largest family (Siezen and Leunissen, 1997). They have the same active site geometry and the catalysis proceeds via the same mechanism but the general three-dimensional structure is different in the two families. Three residues (histidine, aspartic acid and serine) that form the catalytic triad are essential in the catalytic process of the enzyme.



Plant proteases such as papain and bromelain are included in the cysteine proteases family. The essential cysteine and histidine residues play the same role as serine and histidine residues in serine proteases during the catalytic process. Most of aspartic proteases belong to the pepsin family. The pepsin family includes digestive enzymes such as pepsin and chymosin, fungal proteases and processing enzymes such as renin. A second family comprises viral proteases such as the proteases from the HIV also called retropepsin. Metallo proteases are found in bacteria, fungi as well as in higher organisms. They differ widely in their sequences but the great majority of enzymes contain a zinc atom which is catalytically active. In some cases, zinc may be replaced by another metal such as cobalt or nickel without loss of activity (Rao *et al.*, 1998).

#### 2.1.1 Subtilisin-like serine protease

Subtilisin-like serine proteases (EC 3.4.21.62), termed “subtilases” occur in archaea, bacteria, fungi, yeasts and higher eukaryotes. Based on sequence homology, a subdivision into six families is proposed (Siezen and Leunissen, 1997). They are subtilisin family, thermitase, proteinase K, lantibiotic peptidase, kexin and pyrolysins family. Subtilisins are enzymes that are secreted by members of the genus *Bacillus* with subgroups of true subtilisins, high-alkaline proteases and intracellular proteases. The reactions of subtilisin are hydrolysis of proteins with broad specificity for peptide bonds, and showed a preference for a large uncharged residue in